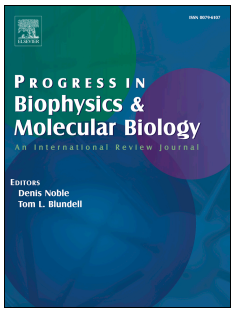
Accepted Manuscript



Promiscuous antibodies characterised by their physico-chemical properties: From sequence to structure and back

Julie M.J. Laffy, Tihomir Dodev, Jamie A. Macpherson, Catherine Townsend, Hui Chun Lu, Deborah Dunn-Walters, Franca Fraternali

PII: S0079-6107(16)30047-5

DOI: 10.1016/j.pbiomolbio.2016.09.002

Reference: JPBM 1144

To appear in: *Progress in Biophysics and Molecular Biology*

Received Date: 14 May 2016

Revised Date: 25 August 2016

Accepted Date: 5 September 2016

Please cite this article as: Laffy, J.M.J., Dodev, T., Macpherson, J.A., Townsend, C., Lu, H.C., Dunn-Walters, D., Fraternali, F., Promiscuous antibodies characterised by their physico-chemical properties: From sequence to structure and back, *Progress in Biophysics and Molecular Biology* (2016), doi: 10.1016/j.pbiomolbio.2016.09.002.

Promiscuous antibodies characterised by their physico-chemical properties: from sequence to structure and back.

Julie MJ Laffy[a]*, Tihomir Dodev[b]*, Jamie A Macpherson[a] , Catherine Townsend[b], Hui Chun Lu[a] , Deborah Dunn-Walters[b,c] & Franca Fraternali[a+]

ABSTRACT

Human B cells produce antibodies, which bind to their cognate antigen based on distinct molecular properties of the antibody CDR loop. We have analysed a set of 10 antibodies showing a clear difference in their binding properties to a panel of antigens, resulting in two subsets of antibodies with a distinct binding phenotype. We call the observed binding multiplicity 'promiscuous' and selected physico-chemical CDRH3 characteristics and conformational preferences may characterise these promiscuous antibodies. To classify CDRH3 physico-chemical properties playing a role in their binding properties, we used statistical analyses of the sequences annotated by Kidera factors. To characterise structure-function requirements for antigen binding multiplicity we employed Molecular Modelling and Monte Carlo based coarse-grained simulations. The ability to predict the molecular causes of promiscuous, multi-binding behaviour would greatly improve the efficiency of the therapeutic antibody discovery process.

* These authors contributed equally to the work
[+] Corresponding author
[a] Randall Division of Cell and Molecular Biophysics, King's College London UK
[b] Department of Immunobiology, King's College London UK
[c] Faculty of Health and Medical Sciences, University of Surrey UK

Keywords: Antibody CDRH3, binding promiscuity, conformational preferences, ELISA, Kidera Factors, Molecular Modelling, Monte Carlo simulations.

INTRODUCTION

Immunoglobulins (Ig) are a crucial component of the adaptive immune response. Adaptive immunity is distinct from innate immunity in that it confers a highly specific defense against invading pathogens and is capable of creating memory against foreign molecules (antigens), enabling a rapid response upon repeated exposure to the same antigen. Immunoglobulins are produced by B cells and are either displayed on the cell surface, as B cell receptors, or are secreted into the extracellular environment and circulate as antibodies in the blood.

Antigen recognition is mediated by the antibody variable regions, which are located at each of the two apical sites on the 'Y' arms of the antibody. A huge diversity of specificities in the antibody repertoire is achieved by gene rearrangement processes, whereby variable (V), diversity (D) and Joining (J) genes recombine to produce a complete heavy chain (VDJ) or light chain (VJ) variable gene. The rearranged gene is expressed in conjunction with a constant region that confers the functional attributes of the antibody. The heavy chain and the light chain variable regions of the immunoglobulin protein fold to form a conserved β-sheet framework interspersed by six hypervariable loops or "complementarity determining regions" (CDRs), so-called because they come together to form the antigen-binding site. Amongst the CDRs (there are three from each chain), the third loop on the heavy chain (CDR-H3) is the most diverse because it is encoded by a stretch of nucleotides spanning all three IGHV-D-J gene segments. Similarly, the equivalent light chain region (CDR-L3) is also diverse although not to the same extent. Site-directed mutagenesis and loop grafting studies have shown that the H3 loop can be sufficient to define antibody

specificity (1). Crystal structure analyses have also identified the CDR-H3 region as being centrally positioned in the antigen-binding site, always in contact with antigen and, in some cases, able to change conformation upon binding (2). In fact, CDR-H3 is the only exception to the canonical structure model, which has identified all remaining CDRs as belonging to one of a few discrete conformations on the basis of sequence length and composition (3-6).

A consequence of the random nature of the Ig gene rearrangement process is that a significant proportion of antibodies produced in the bone marrow may be autoreactive (7). Potentially dangerous autoreactive antibodies must be removed from the repertoire at tolerance checkpoints in the development process in order to avoid autoimmune diseases such as systemic lupus erythematosus (8) and rheumatoid arthritis (9). Some antibodies are capable of binding multiple chemically and structurally diverse antigens, which means that although an antibody may be produced with the potential to usefully bind to exogenous antigens, it may result in binding to self-antigens. Thus previous literature on tolerance and autoimmunity will quite often refer to "polyreactivity" of an antibody. Polyreactive antibodies occur in normal human sera and are thought to act as a first line of defense against foreign antigens (10). They have been shown to cause bacterial lysis (11, 12), induce complement and clear apoptotic cells (13). So in the case of polyreactive antibodies a trade-off balance between potentially useful initial activity and potentially harmful anti-self effects has to be maintained.

In the literature people refer indiscriminately to polyreactive and/or polyspecific antibodies and the definition intrinsic in the prefix 'poly' is matter of debate (14).

It has been recently clarified that this term does not refer to the case of multiple binding due to some stickiness of the antibody chemico-physical properties and that a large screen over a panel of putative antigens is needed before extracting antibodies that clearly react specifically to a subset of these target antigens. Another term that is frequently used when referring to antibodies multi-binding behavior is promiscuity (15, 16). This term has long been investigated in the field of enzyme binding (17), and has been referred to as functional promiscuity, again implying some specific functional behavior resulting from the particular type of binding mode. It is claimed by Favia at al. (17) that the term promiscuous may imply at times what they refer to as 'invisible' phenotypes, observable only under certain conditions. In the past our laboratory has adopted the term promiscuity as generally applicable to proteins that show multiple partners but may use different strategies to bind in such a polyvalent way (18). We use herewith the term promiscuous in referring to a subset of antibodies where binding to a number of tested antigens in the same experimental conditions were detected, as compared to others that do not show binding.

Despite its importance in human health and disease, the molecular basis of antibody multi-binding behaviour remains largely obscure. Towards this goal, it has been noted that some polyreactive antibodies have particularly high frequencies of aromatic amino acids in the CDR-H3 region (19) and high isoelectric points in heavy chain Fv regions. It has also been hypothesised that exposed hydrophobic patches are associated with so-called antibody promiscuity (20). Somewhat paradoxically however, a role for specific hydrogen bonding has also been proposed (21). Comparisons of germline versus antigen-induced antibodies have shown that the former are more likely to be

polyreactive (22) and more flexible (23), which may suggest that one of the features of polyreactive antibodies is enhanced flexibility. In support of this, a structural analysis of a promiscuous antibody in the free and bound states has demonstrated that it adopts a different conformation when bound (24). In contrast however, Sethi et al. (25) have more recently shown that structurally diverse epitopes (the precise binding site on the antigen) bind differentially to a structurally common paratope (the precise binding site on the antibody), implying that paratope flexibility can be limited.

Here, we compare the structural properties of ten antibodies, four of which showed a promiscuous phenotype in enzyme linked immunosorbent assay (ELISA) and six that did not show any binding to the antigens tested. Since the encoding sequences had quite similar characteristics, we wanted to investigate the nature of the promiscuous phenotype and whether the structural properties of the antibody recognition regions were very different in these two datasets.

Using computational methods of Molecular Modelling, Kidera factor clustering and Monte Carlo conformational sampling, we identified physico-chemical properties of the CDR-H3 that are capable of distinguishing between the two phenotypes of antibody presented in this study.

MATERIALS AND METHODS

Identification of antibody heavy and light chain variable region sequences for cloning

The heavy and light chain variable region sequences for the ten antibodies used in this study are shown in Supplementary Materials Table S1. The heavy and light chain sequences had been selected, as part of another study on ageing, from datasets containing unpaired heavy and light chain variable region sequences. The datasets from which the sequences were selected are available at http://www.bcell.org.uk. As there was no information on the natural heavy-light chain pairings, the selected pairing was guided by known paired sequences in the literature (26).

Synthesis of antibodies

The chosen antibody heavy and light chain variable region nucleotide sequences were synthesised by Integrated DNA Technologies (Iowa, USA). The variable region sequences were cloned into heavy (encoding human IgG1 constant region) or light (encoding human kappa or lambda constant region) chain expression vectors using Polymerase Incomplete Primer Extension (PIPE) cloning as described previously (27). Plasmids were then transfected into Adherent FreeStyleTM 293-F cells (Life Technologies) using FuGENE HD (Promega) according to the manufacturer's instructions. The transfected cells were maintained at 70-80% confluence at 37^{o}C, 5% CO_2 in DMEM GlutaMax (31966; Gibco) supplemented with 10% (v/v) fetal calf serum (FCS) (Gibco), penicillin (5000 U/mL, Life Technologies), streptomycin (100 µg/mL, Life Technologies) and hygromycin-B (50 µg/ml, Life Technologies). Antibody-containing tissue culture supernatants were collected seven days post-

transfection. The supernatants were centrifuged at 1000 g for 15 minutes, filtered (0.22μm filters, Sartorius) and stored at 4°C until use. Antibody concentration was determined by ELISA using a standard curve generated from an antibody of known concentration, as set out in Tiller et al (2008) (28) .

ELISA

Antibodies in tissue culture supernatant were diluted to 1μg/ml in PBS and were tested in triplicate against the following antigens: dsDNA (D1501, Sigma), lipopolysaccharide (LPS) (L7770, Sigma), insulin (91077C, SAFC Biosciences) and HEp-2 cell lysate (ABIN964023, Antibodies-online). Antibodies were also tested in wells that had been coated with PBS only. The choice of antigens was based on those commonly used in the literature (7, 28).

The ELISA protocol was as set out in Tiller et al (2008) (28) except for the assay development. Assays were developed by adding 50μl OPD substrate (5mg OPD tablet (P6912-50 TAB, Sigma) dissolved in 10ml 1X peroxide substrate buffer (34062, Thermo Scientific) and incubating in the dark for 30 minutes. The reaction was stopped using 50μl 3M HCl. The optical density was read at 492nm (OD_{492}). A result was deemed positive if the mean OD_{492} was significantly higher than the background (i.e. secondary antibody only).

Antibody structures

Two structural datasets were built to analyse the distribution of CDR-H3 lengths in antibodies from all species and from humans, respectively. The "Non-redundant CDR search" option in SAbDab (Structural Antibody Database;

accessible at <http://opig.stats.ox.ac.uk/webapps/sabdab>) (29) was used to include only complete CDR-H3 regions from crystal structures solved at a resolution of ≤ 3Å. CDR-H3 lengths were determined according to the Chothia definition. The PDB codes for both datasets are available in the supplementary material (Supplementary Materials Table S2).

Antibody Modelling

The modelling workflow is described in Fig. 1 and the steps are outlined below.

[Fig. 1. Modelling workflow]

Step 1: Template Search

The search for templates was carried out using the PIGS web server (Prediction of ImmunoGlobulin Structure; available at <http://www.biocomputing.it/pigs>) (30, 31) and MOE (Molecular Operating Environment; Version 2013.08; Chemical Computing Group, Montreal, Quebec, Canada). Both programs are considered to be state of the art in antibody modelling but differ in their methodologies (32, 33) PIGS employs the "canonical structure" (CS) approach, which rests on knowledge that five out of the six CDRs adopt a limited number of conformations (so-called canonical structures) whose sequence determinants have been identified (4, 5). The exception to this, however, is the CDR-H3, whose sequence length and composition have so far proven too variable for complete CS rules to be identified (3, 6, 34-36). Indeed, matching CDR-H3 conformations becomes increasingly difficult for cases in which the loop is particularly long. In some cases, PIGS was found to perform less when CDR-H3 sequences were

queried, as previously pointed out (30, 31, 37). Furthermore, and most important in our case, PIGS does not allow the user to search for templates specific to one CDR region. For these reasons, we used instead the Antibody Homology Modelling application suite in MOE to search for a CDR-H3 specific template.

Step 2: HMM Multiple Sequence Alignment

The template and query sequences provide the input for Step 2: HMM Multiple Sequence Alignment. Heavy chain (IgH) and light chain (Igκ/Igλ) alignments are generated independently using isotype-specific HMM profiles developed by Tramontano and coworkers, as described in (31). Using this alignment method, gaps in the sequence are added outwards from the centre of the CDRs and conserved residues are always aligned.

Step 3: Model Building and Selection

Step 3 requires the combined H-L alignment from Step 2 and the template structures retrieved in Step 1. 200 structural models are built using the comparative modelling procedure in MODELLER (38) and the best model is selected using normalised DOPE scoring as the quality measure. Models with scores above -1 indicate the presence of nonnative interactions and were not considered. Models were visualised with Visual Molecular Dynamics (VMD; <http://www.ks.uiuc.edu/>) (39).

Step 4: Conformational Sampling

In the last step, the previously selected best model is energy-minimised using the OPLS-AA force field (40) and then used as the starting structure for tCONCOORD simulations (41, 42). From this, an ensemble of 500 tCONCOORD structures is generated using default parameters. We used this procedure, analogously to our previous work (18) because this method efficiently samples the conformational space (see Figure S6 for root-mean square distribution) and is therefore suited to antibody loop sampling, otherwise very difficult to tackle with molecular dynamics or any other atomistic-based sampling. This method is not affected by convergence issues, and has been demonstrated to extract representatives of the structural variability of proteins on the basis of both MD simulations and experimental data (18).

In essence, tCONCOORD samples alternative conformational states *in-vacuo* by fulfilling a set of geometrical constraints extracted from the initial coordinates and interaction types (e.g., covalent bonds, hydrogen bonds, salt bridges, or hydrophobic interactions).

The tCONCOORD constraint definition is based on a statistical analysis of high-resolution X-ray structures and includes a 'solvation score' measure of hydrogen bond stability (41). Based on this description the structure of the protein is rebuilt many hundreds of times, leading to an ensemble that can be analysed.

Clustering analysis

A hierarchical clustering method was employed using the hclust function in R (v3.2.3) and a Minkowski distance, p, of 4. Antibodies were clustered according to the Kidera factors of their CDR-3 region amino acids (Chothia definition). The Kidera factors have been derived to encode 188 physical properties of the 20 amino

acids using dimension reduction techniques. They consist of a 10-dimensional vector of orthogonal factors for each amino acid (43).

Secondary structure analysis

Models were first protonated using MOE. Secondary structure predictions were performed with DSSP (44) and the output was recorded as follows: extended β-strands and β-bridges as "Beta"; α-helices, 3_{10}-helices and π-helices as "Helix"; 3, 4, and 5 turns and non-hydrogen bonded bends as "Turn"; random coil as "Coil". Normalised secondary structure probabilities were calculated for the CDR-H3 and CDR-L3 (Chothia definition).

The 95% confidence intervals were calculated using a bootstrapping method (*boot()* function in R) to generate 100 randomly resampled subsets for each of the reference datasets. This allowed any biases resulting from dataset selection to be avoided.

Conformational analysis

The maximally correlated motion or MCM has been calculated for the tCONCOORD trajectories. This is usually expressed as a linear combination of principal components (PCs) derived from a principal component analysis (PCA) of the system trajectory. We considered here only the first 2 PCA eigenvector components of the trajectory as these are contributing mostly to the total Eigenvalue displacement (in nm^2) (Supplementary Materials Fig. S3). To identify the CDR-3 atoms collective motion, which maximally correlate with these two

PCA components we extracted the projection of the CDR-3 atoms trajectory onto the 1st and the 2nd PCs.

Models are available upon request to the authors' e-mails:

Julie MJ Laffy: julie.laffy@kcl.ac.uk;

Franca Fraternali: franca.fraternali@kcl.ac.uk.

RESULTS AND DISCUSSION

Identification of promiscuous antibodies

Ten antibodies were tested in an ELISA against a panel of four different antigens and blank, uncoated wells. It was found that the ten antibodies could be classed as showing one of the two following phenotypes: (i) a negative signal in all the ELISA wells (GF2, GF3, GF4, GF6, GF8, GF10 in Fig. 2) and (ii) a positive signal in all the ELISA wells (GF1, GF5, GF7, GF9 in Fig. 2). We named the antibodies which showed a positive signal under every ELISA condition in which they were tested "promiscuous". Multiple sequence alignments of the promiscuous and non-promiscuous antibodies are shown in Fig. S1.

[Figure 2. Promiscuous antibodies identified by ELISA]

We note that the promiscuous antibodies showed a positive signal even in the absence of antigen on the ELISA plate. The reactive nature of a promiscuous class of antibodies could be such that one might expect their binding to empty wells in addition to other antigens. Nevertheless, the promiscuous antibodies clearly possess a different phenotype to the non-promiscuous antibodies, which can be detected in this ELISA format, and are relevant to previous literature where these methods have been extensively used to assess polyreactivity/autoreactivity (7, 28, 45). We therefore investigated this phenotype further using computational methods.

CDR-H3 loop length and sequence composition

Numerous previous studies have shown that the CDR-H3 loop is much more variable in its sequence length compared to that of the remaining CDRs, and that this is a critical determinant of its conformation. Fig. 3 shows the distribution of CDR-H3 lengths of the ten antibodies investigated in this paper (pink), a human antibody repertoire (blue) (data from Wu et al. (46)) and antibody structures deposited in the PDB (orange, human; green, all species).

[Figure 3: CDR-H3 length distributions]

The CDR-H3 sequence length distribution of the GF antibodies analysed in this study were found to be longer, on average, when compared to both the human sequence repertoire and the PDB. Previous work has associated longer CDR-H3 regions with antibody binding multiplicity, which may render this observation

significant [ref]. It should be noted however that there is no significant difference in length between the promiscuous and non-promiscuous GF subsets analysed here.

The PDB length distributions displayed a strong preference for shorter CDR-H3 loops. This may be due to the complications associated with structure determination of long, flexible regions of polypeptide. In any case, the discrepancy in CDR-H3 length between the GF antibodies and the PDB structures meant that our template search space was confined to the subset of adequate CDR-H3 lengths illustrated by the shaded green region in Figure 3.

[Figure 3. Distributions of CDR-H3 lengths in the GF antibodies (pink), the human antibody repertoire (blue) and the Protein Data Bank (green and orange).]

Kidera factors and β-sheet propensities are good indicators of GF promiscuity

Attempts to classify promiscuous antibodies from sequence information alone have generally been unsuccessful (see review by (14)). Although some studies have identified the CDR-H3 region as being accountable for the promiscuous nature of antibodies (47-49), this region is always critical in antigen recognition, regardless of promiscuous behaviour.

With this in mind, we first sought to quantify the sequence-based similarities amongst the GF antibodies using a hierarchical clustering method (see Materials

and Methods) and postulated that the inclusion of implicit structural information might improve the quality of our clustering analysis. As such, for each of the ten GF antibodies, protein sequences within the CDR-H3 and -L3 loops were assigned Kidera factors (a set of scores, which quantify physicochemical properties of protein sequences (see Materials and Methods) (43).

The CDR-L3 loop was considered in this analysis, in addition to the –H3 loop, because of its demonstrable function in conferring antibody-antigen specificity. The assignment of Kidera factors to the two sequence regions were combined to give a score for each of the ten antibodies, providing a set of scores by which to cluster.

[Figure 4. Promiscuous (red) and non-promiscuous (blue) GF antibodies can be distinguished on the basis of (A) CDR-3 Kidera properties and (B) CDR-H3 β-sheet occupancy.]

Three principal clusters were defined (Fig. 4A) from this clustering procedure using Minkowski distance as the metric (see Materials and Methods). Of these clusters, two were comprised exclusively of either promiscuous or non-promiscuous antibodies. More specifically, the largest cluster (Cluster 1) contained 66.7% of the total number of non-promiscuous antibodies and 0% of the promiscuous antibodies while Cluster 3 contained 75% of the promiscuous antibodies and 0% of the non-promiscuous antibodies. Cluster 2, on the other hand, contained a mixture of both classes (25% of promiscuous cases and 33.3%

of non-promiscuous cases). Promiscuous GF1 and GF7 and non-promiscuous GF2 and GF4 were the most rooted in their respective clusters, which suggested that the traits associated with either binding phenotype should be best represented by these antibodies.

It is worth noting that similar clustering analyses were performed using simpler parameters to describe the CDR-3 regions, such as hydrophobicity and charge, but no clear separation between the two sets was observed (data not shown). It can therefore be concluded that Kidera Factors are a powerful means of incorporating general physico-chemical properties and structural information into a sequence-based analysis. Interestingly, in the context of blood repertoire, similar conclusions were reached in a study by Epstein et al. (50), which used Kidera factors to analyse TCR repertoires from different individuals.

Enriched β-sheet content in promiscuous CDR-H3 regions

The variable regions of each of the GF antibodies were modelled to look for explicit structural differences between members of the different clusters (see Fig. 4). In each case, the most native-like representative from a sample of 200 models was selected (presented in Fig. 4B) (see Materials and Methods for more information).

Intriguingly, the β-sheet content of the CDR-H3 was highly enriched in the promiscuous antibody models, particularly in those within Cluster 3 (Fig. 4). Moreover, the only promiscuous antibody not belonging to Cluster 3 (GF9) was

also the only one that appeared to lack the ordered β-sheet seen in the other cases. In fact, the CDR-H3 secondary structure of GF9 much more closely resembled a random coil; a feature that was shared amongst all antibodies in Clusters 1 and 2.

The β-sheet-forming propensities (51) of residues in the CDR-H3 regions (Fig. 5) also reflected the results in Fig. 4. Within Cluster 3, β-sheet breakers (pink) and strong β-sheet breakers (red) were concentrated at the apices of the CDR-H3s and were largely absent from their stems, thereby permitting anti-parallel β-sheet hydrogen bonding. The exception to this was a serine residue in GF5 (breaker; pink), however its geometry and neighbouring environment was such that a hydrogen bond still formed. Contrastingly, the distributions of β-sheet breakers were far more scattered in members of Clusters 1 and 2. A higher occurrence of strong β-sheet formers (green) in the stems of Cluster 3 members was also noted. Once again, this observation was less relevant in GF5. To this point, it is perhaps worth noting that GF5 was the least deep-seated in Cluster 3 (Fig. 4).

[Figure 5. β-sheet propensities and structure of CDR-H3 residues.]

Beyond β-sheet propensities, the degree of CDR-H3 collapse seemed to correlate well with membership to a particular cluster (Fig. 5). Cluster 1 CDR-H3 members were the most ‘tangled’; Cluster 2 members were comparatively taut and Cluster 3 members were most extended (forming β-sheet structure).

To assess the statistical significance of this finding and to ensure that differences in secondary structure content were not simply a reflection of the templates used, Monte Carlo-based simulations were run on all GF antibodies using the tCONCOORD program (see Materials and Methods). CDR-H3 secondary structure probabilities were calculated for the GF ensembles of 500 tCONCOORD structures and subsequently averaged over the promiscuous and non-promiscuous sets.

[Figure 6. Average secondary structure probabilities in simulation ensembles for (i) promiscuous and (ii) non-promiscuous antibodies.]

As illustrated in Fig. 5, the CDR-H3 β-sheet content was 33.4% higher in the promiscuous antibodies compared to the non-promiscuous antibodies (two-tailed P value < 0.0001) and this difference was even more pronounced when GF9 was excluded from the promiscuous set (data not shown). The non-promiscuous antibodies were also predicted to have 18% more 'turn' secondary structure (defined as 3, 4, and 5 turns and non-hydrogen bonded bends) (two-tailed P value < 0.0001).

Predicted secondary structure occupancies were also calculated for the CDR-L3 regions of the promiscuous and non-promiscuous antibodies however no significant differences were observed (Supplementary Materials Fig. S2).

Key residues in the CDR-H3 of GF1

To reduce the high dimensionality of the tCONCOORD trajectories and to identify the dominant molecular motions related to either promiscuous or non-

promiscuous antibodies, principal component analysis (PCA) was applied. Two representative tCONCOORD trajectories were taken from each class; GF1 for the promiscuous subset and GF4 for the non-promiscuous subset. As mentioned earlier in the text, these antibodies were amongst the most deep-rooted in their respective clusters (Fig. 4A). Furthermore, their CDR-H3 and CDR-L3 loops were determined to be of approximately the same length and of very high sequence similarity (CDR-H3, 99%; CDR-L3, 95%). In the trajectories of both GF1 and GF4 the dominant global motion was found to be a side-to-side 'rocking' of the heavy and light chain framework regions (Supplementary Materials Fig. S4). When the projection of atoms in the CDR-H3 and -L3 loops onto PC1 and PC2 were determined, several residues in the CDR-H3 loops of the promiscuous GF1 antibody, but not GF4, were also found to contribute to the global motion accounted for by the first and second eigenvectors (Fig. 7). The CDR-L3 regions were found not to contribute significantly to the global motion in either the first or the second eigenvectors (Supplementary Materials Fig. S5).

[Figure 7. Principal component analysis of CDR-H3 dynamics.]

Analysis of PC1 and PC2 highlighted four particularly dynamic residues in the CDR-H3 of GF1: Lys100, Asp100B, His100C and Trp100E (Chothia numbering definition; Fig. 7A). Interestingly, these residues are concentrated at the apex of the CDR-H3 β-sheet (Fig. 7B). Furthermore, some of these residues contain features that are indicative of binding versatility. Firstly, residues such as Lys and Trp are often found at binding interfaces due to their ability to adapt to both hydrophobic and hydrophilic environments (52). Residues with a large surface

area (Lys, His, Trp) should also contribute favourably to binding free energies with antigen due to their participation in van der Waals' interactions (4, 5, 53). Lastly, some residues, including Tyr and Trp, are intrinsically flexible (54); a property that has previously been found as characteristic of promiscuous residues in hub proteins (18). Such residues could sample a wider range of dihedral angles and adjust their fit to accommodate different antigens.

Putative binding mechanisms

We hypothesize that the advantage of a protruding β-sheet-rich CDR-H3 in the promiscuous antibodies lies in the inability of its residues to partake in intra-molecular interactions. Instead, such outwardly positioned residues would be most effectively placed to encourage interactions with foreign molecules. In contrast, the lower occurrence of β-sheet in the non-promiscuous CDR-H3 loops would result in a higher number of 'rigidifying' intra-molecular CDR-H3 contacts. This could explain why the CDR-H3 residues in GF4 were less dynamic than their counterparts in GF1 despite residing in an otherwise flexible random coil (Fig. 7).

Beyond this, the findings presented here suggest that specific residues in the GF antibodies confer promiscuity, as was found in a large-scale analysis of multi-binding proteins (18). The dominant nature of the promiscuous CDR-H3 could permit the exposure of local and chemically versatile neighbourhoods of amino acids with configurations free of the steric or structural constraints normally expected within the antibody-combining site.

Conclusions

Binding promiscuity is a characteristic of many proteins and is an essential feature for efficient cellular communication. In the past (18) we characterised the essential role played by promiscuous residues at the surface of multi-binding proteins (hubs) and revealed an important feature the flexibility exerted by these residues and their enhanced motion. This plasticity can be efficiently exploited in supporting diversity of interaction and effective binding. Here we focused on another important class of proteins, antibodies, and highlighted the important role of physico-chemical properties of specific residues at the CDR-H3 in conferring promiscuous behaviour to a set of antibodies tested for multiple binding to a small panel of antigens *in-vitro*. Interestingly, we found that it is not the entire region of the paratope that has to exhibit flexibility to effectively modulate promiscuous binding, but this comes at a compromise between the β-sheet content of the CDR-H3 that can hold in place and protrude out of the antibody combining-site, thereby 'offering' to the epitope the crucially binding effective residues that can exert their plastic role and explore multiple binding conditions. In analogy with the previous findings, the promiscuity is relegated to only few versatile and intrinsically flexible residues. These are concentrated at the apex of the CDR-H3 β-sheet and found to contribute to the global motion accounted for by the first and second eigenvectors of the variable framework region. The observed properties were crucially different from the ones measured for a comparable class of antibodies that showed no binding preferences in the same tested conditions.

The findings presented here can have important applications in the design of specific antibodies devoid of promiscuous behaviour, as it is well known that in many circumstances antibodies elicited towards a specific antigen are found to

bind to structurally unrelated epitopes (55). Additionally, the role of specific physicochemical determinants in promiscuous protein activities could pave the way to alternative approaches to directed evolution of novel protein functions (56).

Acknowledgements

The authors would like to acknowledge the help of Jens Kleinjung for setting up and maintaining the computer clusters on which the calculations in this paper were performed. Bryan (Yu-Chang) Wu, David Kipling and Joselli Silva O'Hare for sequence-data support. FF and DDW acknowledge funding from the MRC/BBSRC programme grant MR/L01257X/1. TD acknowledges the Dunhill Medical Trust (grant number R279/0213). CT acknowledges the BBSRC CASE award (number BB/L015845/1).

References

1. Xu JL, Davis MM. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. Immunity. 2000;13(1):37-45.
2. Stanfield RL, Wilson IA. Antigen-induced conformational changes in antibodies: a problem for structural prediction and design. Trends Biotechnol. 1994;12(7):275-9.
3. Shirai H, Kidera A, Nakamura H. Structural classification of CDR-H3 in antibodies. FEBS letters. 1996;399(1-2):1-8.
4. Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. J Mol Biol. 1987;196(4):901-17.
5. Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, et al. Conformations of immunoglobulin hypervariable regions. Nature. 1989;342(6252):877-83.
6. Shirai H, Kidera A, Nakamura H. H3-rules: identification of CDR-H3 structures in antibodies. FEBS letters. 1999;455(1-2):188-97.
7. Wardemann H, Yurasov S, Schaefer A, Young JW, Meffre E, Nussenzweig MC. Predominant autoantibody production by early human B cell precursors. Science. 2003;301(5638):1374-7.
8. Yurasov S, Wardemann H, Hammersen J, Tsuiji M, Meffre E, Pascual V, et al. Defective B cell tolerance checkpoints in systemic lupus erythematosus. J Exp Med. 2005;201(5):703-11.
9. Samuels J, Ng YS, Coupillaud C, Paget D, Meffre E. Impaired early B cell tolerance in patients with rheumatoid arthritis. J Exp Med. 2005;201(10):1659-67.
10. Guilbert B, Dighiero G, Avrameas S. Naturally occurring antibodies against nine common antigens in human sera. I. Detection, isolation and characterization. J Immunol. 1982;128(6):2779-87.
11. Zhou Z-H, Tzioufas AG, Notkins AL. Properties and function of polyreactive antibodies and polyreactive antigen-binding B cells. Journal of autoimmunity. 2007;29(4):219-28.
12. Zhou ZH, Zhang Y, Hu YF, Wahl LM, Cisar JO, Notkins AL. The broad antibacterial activity of the natural antibody repertoire is due to polyreactive antibodies. Cell Host Microbe. 2007;1(1):51-61.
13. Zhou ZH, Wild T, Xiong Y, Sylvers P, Zhang Y, Zhang L, et al. Polyreactive antibodies plus complement enhance the phagocytosis of cells made apoptotic by UV-light or HIV. Sci Rep. 2013;3:2271.
14. Dimitrov JD, Planchais C, Roumenina LT, Vassilev TL, Kaveri SV, Lacroix-Desmazes S. Antibody Polyreactivity in Health and Disease: Statu Variabilis. The Journal of Immunology. 2013;191(3):993-9.
15. James LC, Tawfik DS. The specificity of cross-reactivity: Promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness. Protein Science. 2003;12(10):2183-93.
16. James LC, Tawfik DS. Conformational diversity and protein evolution - a 60-year-old hypothesis revisited. Trends in Biochemical Sciences. 2003;28(7):361-8.
17. Nobeli I, Favia AD, Thornton JM. Protein promiscuity and its implications for biotechnology. Nature Biotechnology. 2009;27(2):157-67.

18. Fornili A, Pandini A, Lu HC, Fraternali F. Specialized Dynamical Properties of Promiscuous Residues Revealed by Simulated Conformational Ensembles. J Chem Theory Comput. 2013;9(11):5127-47.
19. Droupadi PR, Varga JM, Linthicum DS. Mechanism of allergenic cross-reactions--IV. Evidence for participation of aromatic residues in the ligand binding site of two multi-specific IgE monoclonal antibodies. Mol Immunol. 1994;31(7):537-48.
20. Barbas CF, 3rd, Heine A, Zhong G, Hoffmann T, Gramatikova S, Bjornestedt R, et al. Immune versus natural selection: antibody aldolases with enzymic rates but broader scope. Science. 1997;278(5346):2085-92.
21. James LC, Tawfik DS. The specificity of cross-reactivity: promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness. Protein Sci. 2003;12(10):2183-93.
22. Chen C, Stenzel-Poore MP, Rittenberg MB. Natural auto- and polyreactive antibodies differing from antigen-induced antibodies in the H chain CDR3. The Journal of Immunology. 1991;147(7):2359-67.
23. Manivel V, Sahoo NC, Salunke DM, Rao KV. Maturation of an antibody response is governed by modulations in flexibility of the antigen-combining site. Immunity. 2000;13(5):611-20.
24. James LC, Roversi P, Tawfik DS. Antibody multispecificity mediated by conformational diversity. Science. 2003;299(5611):1362-7.
25. Sethi DK, Agarwal A, Manivel V, Rao KVS, Salunke DM. Differential epitope positioning within the germline antibody paratope enhances promiscuity in the primary immune response. Immunity. 2006;24(4):429-38.
26. DeKosky BJ, Kojima T, Rodin A, Charab W, Ippolito GC, Ellington AD, et al. In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. Nat Med. 2015;21(1):86-91.
27. Dodev TS, Karagiannis P, Gilbert AE, Josephs DH, Bowen H, James LK, et al. A tool kit for rapid cloning and expression of recombinant antibodies. Sci Rep. 2014;4:5885.
28. Tiller T, Meffre E, Yurasov S, Tsuiji M, Nussenzweig MC, Wardemann H. Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. Journal of Immunological Methods. 2008;329(1-2):112-24.
29. Dunbar J, Krawczyk K, Leem J, Baker T, Fuchs A, Georges G, et al. SAbDab: the structural antibody database. Nucleic Acids Research. 2014;42(Database issue):D1140-6.
30. Marcatili P, Rosi A, Tramontano A. PIGS: automatic prediction of antibody structures. Bioinformatics (Oxford, England). 2008;24(17):1953-4.
31. Marcatili P, Olimpieri PP, Chailyan A, Tramontano A. Antibody structural modeling with prediction of immunoglobulin structure (PIGS). Nature protocols. 2014;9(12):2771-83.
32. Almagro JC, Teplyakov A, Luo J, Sweet RW, Kodangattil S, Hernandez-Guzman F, et al. Second antibody modeling assessment (AMA-II). Proteins. 2014;82(8):1553-62.
33. Almagro JC, Beavers MP, Hernandez-Guzman F, Maier J, Shaulsky J, Butenhof K, et al. Antibody modeling assessment. Proteins. 2011;79(11):3050-66.
34. Kim ST, Shirai H, Nakajima N, Higo J, Nakamura H. Enhanced conformational diversity search of CDR-H3 in antibodies: role of the first CDR-H3 residue. Proteins. 1999;37(4):683-96.

35. Morea V, Tramontano A, Rustici M, Chothia C, Lesk AM. Conformations of the third hypervariable region in the VH domain of immunoglobulins. J Mol Biol. 1998;275(2):269-94.
36. Shirai H, Nakajima N, Higo J, Kidera A, Nakamura H. Conformational sampling of CDR-H3 in antibodies by multicanonical molecular dynamics simulation. Journal of molecular biology. 1998;278(2):481-96.
37. Marcatili P, Ghiotto F, Tenca C, Chailyan A, Mazzarello AN, Yan XJ, et al. Igs Expressed by Chronic Lymphocytic Leukemia B Cells Show Limited Binding-Site Structure Variability. The Journal of Immunology. 2013;190(11):5771-8.
38. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 1993;234(3):779-815.
39. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14(1):33-8, 27-8.
40. Robertson MJ, Tirado-Rives J, Jorgensen WL. Improved Peptide and Protein Torsional Energetics with the OPLSAA Force Field. J Chem Theory Comput. 2015;11(7):3499-509.
41. Seeliger D, Haas J, de Groot BL. Geometry-based sampling of conformational transitions in proteins. Structure. 2007;15(11):1482-92.
42. Seeliger D, De Groot BL. tCONCOORD-GUI: visually supported conformational sampling of bioactive molecules. J Comput Chem. 2009;30(7):1160-6.
43. Nakai K, Kidera A, Kanehisa M. Cluster analysis of amino acid indices for prediction of protein structure and function. Protein Eng. 1988;2(2):93-100.
44. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22(12):2577-637.
45. Meffre E, Schaefer A, Wardemann H, Wilson P, Davis E, Nussenzweig MC. Surrogate light chain expressing human peripheral B cells produce self-reactive antibodies. Journal of Experimental Medicine. 2004;199(1):145-50.
46. Wu YC, Kipling D, Dunn-Walters DK. Age-Related Changes in Human Peripheral Blood IGH Repertoire Following Vaccination. Front Immunol. 2012;3:193.
47. Ditzel HJ, Itoh K, Burton DR. Determinants of polyreactivity in a large panel of recombinant human antibodies from HIV-1 infection. J Immunol. 1996;157(2):739-49.
48. Martin T, Crouzier R, Weber JC, Kipps TJ, Pasquali JL. Structure-function studies on a polyreactive (natural) autoantibody. Polyreactivity is dependent on somatically generated sequences in the third complementarity-determining region of the antibody heavy chain. The Journal of Immunology. 1994;152(12):5988-96.
49. Ichiyoshi Y, Casali P. Analysis of the structural correlates for antibody polyreactivity by multiple reassortments of chimeric human immunoglobulin heavy and light chain V segments. Journal of Experimental Medicine. 1994;180(3):885-95.
50. Epstein M, Barenco M, Klein N, Hubank M, Callard RE. Revealing Individual Signatures of Human T Cell CDR3 Sequence Repertoires with Kidera Factors. PLoS ONE. 2014;9(1):e86986-10.
51. G.D. PPJaF. Prediction of protein structure and the principles of protein conformation. Fasman GD, editor. New York: Springer US; 1989.
52. Mian IS, Bradwell AR, Olson AJ. Structure, function and properties of antibody binding sites. J Mol Biol. 1991;217(1):133-51.
53. Gelles J, Klapper MH. Pseudo-dynamic contact surface areas: estimation of apolar bonding. Biochim Biophys Acta. 1978;533(2):465-77.

54. Mian IS, Bradwell AR, Olson AJ. Structure, function and properties of antibody binding sites. Journal of molecular biology. 1991;217(1):133-51.
55. James LC, Tawfik DS. The specificity of cross-reactivity: Promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness. Protein Science. 2009;12(10):2183-93.
56. Buchholz F, Stewart AF. Alteration of Cre recombinase site specificity by substrate-linked protein evolution. Nat Biotechnol. 2001;19(11):1047-52.
57. Chailyan A, Tramontano A, Marcatili P. A database of immunoglobulins with integrated tools: DIGIT. Nucleic Acids Research. 2011;40(D1):D1230-D4.
58. Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang J-M, et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Research. 2011;39(Web Server issue):W13-7.
59. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of molecular biology. 2000;302(1):205-17.

Figure 1. Modelling workflow. The pipeline takes the Fv regions of partner heavy and light chain sequences as input in Step 1 (cyan). The search for templates is conducted using the PIGS web server and the Antibody Homology suite in MOE. A CDR-H3-specific template is extracted from the MOE search. The PIGS "Same antibody" method is used to search for a template whose heavy (H) and light (L) chains are the best combined match for the query H and L sequences. If the sequence identity match of either chain in the PIGS template is below 70%, an additional chain-specific template is selected using the alternative "Best H and L chain" PIGS strategy. Together with the query H and L sequences, the selected template sequences are used as input in Step 2 (green). Heavy chain (IgH) and light chain (Igκ/Igλ) alignments are generated independently using isotype-specific HMM profiles developed by Tramontano and coworkers, as first described in (57). The combined alignment and the template structures identified in Step 1 are used in Step 3 (orange). 200 models are built using the comparative modelling procedure in MODELLER and the best model is selected using normalised DOPE scoring as the quality measure. In the last step (blue), the model is used as the starting structure for tCONCOORD simulations. An ensemble of 500 tCONCOORD structures is generated using default parameters. Pink, input/output; solid red arrow, proceed to the next step in the scheme; dashed red arrow, proceed to a non-immediate step in the scheme.

Figure 2. Promiscuous antibodies identified by ELISA. Ten candidate promiscuous antibodies were screened for binding against DNA, LPS, insulin and HEp2. Four of these antibodies (GF1, GF5, GF7 and GF9) gave positive results in all wells. The remaining six antibodies (GF2, GF3, GF4, GF6, GF8, GF10) produced only negative results. All wells contained phosphate-buffered saline (PBS) solution with Tween (a non-ionic detergent) acting as a blocking agent. The colour key represents the well coating used. No Ab, no antibody; No Ag, no antigen.

Figure 3. CDR-H3 length distributions. Green, antibody PDB structures from all species; Orange, Human antibody PDB structures. Blue, Human sequences from the peripheral blood (data from Wu et al. (46)); Pink, selected GF sequences. The shaded green area represents the range of templates' CDR-H3 lengths used in the

structural modelling of GF antibodies. A CDR-H3 length cutoff threshold of more than 40 amino acids was applied to the PDB set containing all species.

Figure 4. Promiscuous (red) and non-promiscuous (blue) GF antibodies can be distinguished on the basis of (A) CDR-3 Kidera properties and (B) CDR-H3 β-sheet occupancy. (A) Each node on the tree is a GF antibody represented by the Kidera factors of its CDR-3 regions. Data were clustered using the Minkowski metric with a distance *p* of 4. (B) Structural Fv models of the GF antibodies. The CDR-3 regions on the heavy and light chains are highlighted in red and blue for promiscuous and non-promiscuous cases, respectively.

Figure 5. β-sheet propensities and structure of CDR-H3 residues. Backbone representation of the CDR-H3 with amino acids coloured according to their propensity to form β-sheet (51): green, strong formers (V,I,M); blue, formers (F,Y,C,T,W,L,Q); yellow, indifferent (R,G,A,D); pink, breakers (H,S,K,N,P); red, strong breakers (E). Clusters were defined according to the result in Fig. 4. GF labels in red denote promiscuous CDR-H3s; blue labels denote non-promiscuous CDR-H3s.

Figure 6. Average secondary structure probabilities in simulation ensembles for i) promiscuous and ii) non-promiscuous CDR-H3 regions. Predictions were calculated using DSSP. Each antibody is represented by a conformational ensemble of 500 tCONCOORD structures, such that the total number of structures in i) is 2000 (4 antibodies) and in ii) is 3000 (6 antibodies). The error bars represent the confidence intervals at the 95% level, estimated using 100 runs of bootstrap resampling.

Figure 7. Principal component analysis of CDR-H3 dynamics. (A) The contribution of each atom in the CDR-H3 loop of GF1 (red) and GF4 (blue) to the first and second eigenvectors measured in nm^2. (B) The residues whose atoms contributed most significantly to the respective PCs in GF1 were labelled in (A) and mapped onto the GF1 model. Corresponding residues were mapped onto the GF4 model for reference.

Figure S1. Multiple sequence alignments of (A) the heavy chain and (B) the light chain sequences of the promiscuous and non-promiscuous antibodies. Sequences were aligned using the Simple MSA option in T-Coffee (58, 59).

Figure S2. Average secondary structure probabilities in simulation ensembles for i) promiscuous and ii) non-promiscuous CDR-L3 regions. Predictions were calculated using DSSP. Each antibody is represented by a conformational ensemble of 500 tCONCOORD structures, such that the total number of structures in i) is 2000 (4 antibodies) and in ii) is 3000 (6 antibodies).

Figure S3. Eigenvalue spectrum for GF1 (red) and GF4 (blue). The contribution (in nm^2) of the first six eigenvectors to the total global motion of the tCONCOORD ensembles (see Materials and Methods) of GF1 and GF4.

Figure S4. Porcupine representation of the first and second principal components of the promiscuous and non-promiscuous tCONCOORD ensembles. The red spikes represent the direction and relative magnitude of the motion of each C^{α} atom along the first and second principal components.

Figure S5. Principal component analysis of CDR-L3 dynamics. The contribution of each atom in the CDR-L3 loop of GF1 (red) and GF4 (blue) to the first and second eigenvectors measured in nm^2.

Figure S6. Root mean square deviation of CDR-H3 loops from tCONCOORD trajectories. The spatial distribution of the CDR-H3 loops from an ensemble of 500 structures was measured for the promiscuous (red) and non-promiscuous (blue) sets of antibodies.

Table S1. Ten antibody heavy and light chain sequences selected for cloning based on the difference in frequency of their properties.

ID	IGHV	IGHD	IGHJ	CDR-H3	IGLV	IGLJ	CDR-L3
GF1	IGHV3-30	IGHD2-15	IGHJ2	ARVVGSSKWDHAWYDL	IGKV1-37	IGKJ5	QQYNNWPRT
GF2	IGHV4-39	IGHD4-11	IGHJ4	ARGRYRDYSNPACVFDY	IGLV2-8	IGLJ3	SSYAGSNNLRV
GF3	IGHV3-23	IGHD2-15	IGHJ4	AKDWEKYCSGGSCQYDY	IGKV4-1	IGKJ3	QQYYSTPRIT
GF4	IGHV3-30	IGHD5-24	IGHJ4	AKEYLKGRDGYNYYFDY	IGLV2-8	IGLJ3	SSYAGSNNLE
GF5	IGHV3-23	IGHD3-9	IGHJ4	AKASLVRYFDWLFNFDY	IGKV1-27	IGKJ3	QKYNSAPPFT
GF6	IGHV3-33	IGHD2-15	IGHJ4	ARAYNRCSGGSCYEHETTLTGFDY	IGLV1-47	IGLJ3	AAWDDSLSGPSWV
GF7	IGHV3-33	IGHD3-3	IGHJ6	ARDMVLEWSYYYYGMDV	IGKV1D-12	IGKJ2	QQANSFPWT
GF8	IGHV3-30	IGHD3-10	IGHJ6	ASPTPSGSSARYYYYYGMDV	IGLV5-45	IGLJ3	MIWHSSACV
GF9	IGHV3-33	IGHD4-17	IGHJ6	ARGPHRYGDYGGYYYYYGMDV	IGKV1D-12	IGKJ5	QQANSFPIT
GF10	IGHV3-7	IGHD2-21	IGHJ6	AREGCGGDCYSYYYYYYMDV	IGKV1-27	IGKJ1	QVYNSAPPVT

Table S2. PDB codes for all species and Human datasets.

658 Ig structures from all species in the PDB

1iqd 1iqw 4g6f 3fo2 4kvc 3grw 4↵y 3ggw 3zkq 1pz5 1igm 4jr9 4i77 3ifl 1u0q 1nlb 1za6 1q9w 3giz 3fo9 1zan 3kym 3mlr 2osl 4aeh 1opg 1s3k 3k7u 1yej 3inu 4hpy 1dfb 3bn9 1ndg 4hk0 1lo4 1egj 3ux9 3vg9 1sy6 2otu 3eak 2b1h 4dtg 3lrs 3phq 2g5b 3q6g 3sqo 2w60 3tcl 3zkm 1psk 2h9g 1kiq 1mam 4hs8 3m8o 35c8 2vxv 2hwz 3tnn 3utz 3ezj 7fab 1yjd 3efd 4m48 1rih 2cmr 3g5y 1sbs 3k80 3r1g 4jb9 1h8s 3t65 4jo4 4hkx 2zch 3eot 3pp4 1fn4 1ikf 3b9v 1qd0 1fvd 1gig 3r06 1nak 1cly 2qqn 3okd 4ene 2c1p 4hie 2xzc 3go1 2e27 3hc3 43c9 1dl7 4ht1 4at6 3qhz 1uj3 2nr6 1cz8 3t3p 4hc1 4hzl 3l95 3gjf 4jzn 1c5d 4dkf 2adg 2fb4 3auv 1ob1 1yy8 3eba 4ky1 2y6s 1ghf 1f3d 1l7t 4aei 1hil 1a2y 2eh7 4hjj 1igc 2or9 1u6a 3u30 1rhh 1wej 2aju 3ujt 4b41 3nh7 1f8t 2op4 1tqb 1t2j 2x89 3o2d 4g5z 3vfg 4jqi 1ay1 2hfg 1mim 2gsi 4dqo 4fql 1jpt 1mvf 1c5c 4idl 3mck 1nsn 3cfb 4dvb 4jo1 3qxt 1clo 4hxa 3liz 2xzq 3mlw 3tpk 3fzu 1uz8 4fqi 1kcr 1gpo 1ic7 1bz7 4jam 1kb5 4laj 4k3e 2f5a 4↵v 2zpk 2vl5 4amk 2vxq 2fat 1rz7 1mex 3bkj 1ors 1a6v 1sm3 3i02 4gsd 3fn0 3s96 1for 1fh5 3mbx 3pnw 3vi3 1dvf 1ai1 1p7k 1ri8 1n4x 3zkx 4fqq 1kxt 4fz8 2agj 1jfq 3juy 1kxq 2yk1 3g04 3eyf 3b9k 1mnu 3i9g 3u9p 2bdn 1eap 1flr 1mh5 3r0m 4k3j 3qeh 2qqk 1tzi 2q8b 4gag 3dv6 3d85 3gm0 2a9m 3g9a 3s35 4h0i 4jha 2d7t 2ipu 4eig 1mj8 3u7y 4kuc 2z92 3upc 2w9d 1rur 1aqk 1fai 3qpq 3gkw 3sob 2vq1 3na9 2g75 3u1s 1q72 1zea 1frg 1kcv 1ol0 3se8 1xgy 2brr 1igf 4jm4 2o5x 3ifp 4hwe 1jgl 1rjl 1plg 3vw3 1dn0 2jel 3e8u 3lh2 1mqk 3idx 4gmt 2xt1 2g60 3i75 1nj9 1etz 2r0l 3tnm 2hrp 4gq9 1mju 3idg 3d9a 1dzb 1h8n 4jg1 4eow 4jn2 4hs6 4krn 2j4w 2xkn 3qsk 3o6k 3mcl 4fqj 3pgf 2aab 2hh0 3k74 2v7h 3sdy 3b2u 3ijh 3ujj 3ojd 1nbv 1um5 1j05 4hix 1fl5 3uc0 3lex 12e9 1e6o 2xa8 1ind 3u2s 3ls5 1ct8 1jgu 1t4k 1ce1 1jn6 3v6o 1ejo 2uzi 4gft 1jrh 3dvg 3k2u 3mxw 3skj 1n0x 1keg 4j6r 4dka 4leo 2wzp 3iet 3cvi 1qfu 1il1 1nc2 3uji 2vxs 2arj 4fze 4kph 4krm 1mfa 1zv5 1i7z 4ers 3oaz 3ln9 1yc7 4lst 4d9l 1uwe 2xqb 3↵d 1fj1 4f33 1aif 4aq1 3cx5 1pg7 1vge 2dqu 1f4w 4jpk 1kel 4fnl 4gxv 2hmi 3hnt 3ra7 3u0t 4k2u 1cic 3hi6 4jpw 1c12 4m43 1igt 1t3f 3qot 3eo9 2jb5 4k3d 3i2c 4gw4 3so3 3cfd 2xa3 4g3y 4fhb 3dur 2ih3 2bmk 1zvy 4h0h 2aep 2pcp 3ks0 2p45 2cju 4krp 4k7p 1kfa 3rvv 1nl0 2a6i 1ngz 3eo0 1rmf 3stb 4h20 1cr9 3qg6 1ztx 1dee 3bky 3oz9 2zkh 1a6t 1kxv 1fns 2ck0 3h42 3kr3 1hi6 2x1q 3ncj 3o0r 1fe8 1qkz 1ad9 1mvu 1mcp 1q0x 2qhr 3ma9 1tzh 1lk3 2vyr 1yee 1ap2 1baf 3umt 3rkd 1y0l 1f58 1mlb 1dqd 4ebq 4ag4 3uls 2uyl 3h3b 2g2r 1sjv 2yc1 3sge 1a4j 1dlf 3cmo 4b5e 3esu 2ok0 4hxb 1ndm 4imk 3cfi 4lsu 1osp 3nfs 1a3r 4lkc 1hcv 3mo1 2adf 2iq9 2wuc 8fab 4jy5 2gcy 1mrd 1f2x 1e4x 3qyc 1dsf 3c08 3v0w 2h1p 1g9m 2v7n 3ldb 3clf 1jhl 1eo8 1bj1 1fsk 1ncw 3hc4 1kcu 2uud 3s34

1igj 1uwx 2ddq 1lo0 3gk8 4hpo 4f57 1dqq 2j88 3gkz 2xra 3mly 1x9q 4j8r 4hlz 4hcr 3se9 3ntc 3hr5 3o2v 3qcu 3qpx 3q3g 3dif 3hzm 4m61 2hkf 3dsf 1l7i 3lmj 2qsc 1bfv 1t66 4dcq 1ohq 4i9w 4d9q 3v0a 1ngy 4ej1 2ypv 4hbc 3v6f 3u0w 1nld 3esv 4hfw 4jy6 3p9w 2fx7 3h0t 2vxt 1h0d 2r8s 4lmq 1ktr 3ifo 1ggc 2xtj 1dql 4dgy 4g6m 2ny1 2ghw 1emt 3qwo 2gki 1d5i 1ibg 4lsp 3gi9 2x1o 3uyp 3k1k 2xqy 3mj8 1uyw 4fqc 3kdm 3nzh 3nps 1xiw 1i9j 2xwt 3v4u 1c1e 4h88 2oz4 3p0y 1jv5 1gaf 3l5y 3g6a 1wt5 4m1d 1fgn 2r56 1p2c 1oaq 3nz8 1dbb 1p4b 3u6r 1nmb 4jdv 4fq2 3dgg 3ghe 3gnm 1seq 3n9g 2v17 2fbj 2x1p 1bln 1nfd 4al8 1sjx 1ynl 1h3p 1iai 1yqv 2p4a 4jpi 1w72 1nca 2ai0 3gbn 1t2q 3qq9 2aj3 1hq4

248 Ig structures from the Human PDB

1iqd 3u0w 1it9 1n0x 4g6f 4j6r 4fqq 1uwe 3lh2 3grw 2xra 3mbx 3hi6 3hc4 3mlw 2vxs 3fzu 4fqi 3s34 4jpi 1za6 3giz 4gw4 1i7z 3kym 4jam 3mlr 4hpo 2osl 4f57 1opg 2f5a 1s3k 3inu 4hpy 1dfb 1t2j 4d9l 1ad0 2xqb 3bn9 3mly 1x9q 1rz7 4imk 3aaz 3ux9 4hcr 3se9 3ntc 4gsd 3o2v 3fn0 2b1h 4dtg 3lrs 1vge 3qpx 3dvg 1fh5 3qcu 3tcl 3dif 3zkm 4gxv 3upc 1l7i 3nps 3lmj 4ky1 2qsc 3u0t 4fnl 4fz8 3r1g 2agj 3juy 2vxv 1ohq 3tnn 4jpw 1y0l 1ngy 7fab 3g04 3eyf 3qot 3eo9 2jb5 3fo2 4fql 4hfw 2fx7 4k7p 3h0t 1h0d 4lmq 4g3y 2h9g 3uji 2xtj 4jb9 3tnm 3sqo 1igm 1dql 2qqk 3qeh 3p9w 4dgy 4g6m 2ny1 3d85 2xzc 2a9m 4krp 4hkx 4fze 4dkf 3s35 1ikf 1d5i 4jha 2d7t 3b9v 1fvd 3u7y 1ngz 4lsp 3idx 1aqk 4g6a 3gkw 3h42 2qqn 3sob 3na9 2g75 3u1s 3bky 2zkh 4jy6 1ol0 4fqc 4hie 3kdm 3nzh 3qpq 3go1 4hs8 3kr3 3hc3 2xwt 2ghw 4jm4 2o5x 3se8 3uls 1uj3 4hwe 3ma9 3ncj 1dn0 3p0y 1jv5 1gaf 2vyr 4ers 3l5y 3g6a 1wt5 3gjf 4jzn 4m1d 4lst 2aj3 1nl0 2r56 3m8o 3dgg 2hfg 2fb4 3auv 2cmr 3so3 1yy8 3oaz 1dee 3u6r 1cly 3idg 2eiz 4jdv 4fq2 3qhz 1bvk 4eow 1ad9 2yc1 4hjj 3ghe 4hs6 3n9g 4al8 4i77 3mcl 3pgf 1u6a 3u30 1rhh 2hh0 3t2n 3sdy 2uzi 3b2u 3nh7 4lsu 3ujj 3nfs 4d9q 4lkc 1fl5 2x89 3mo1 2yk1 2xa8 3u2s 4g5z 1w72 2wuc 8fab 4jy5 4fqj 3gbn 2vxq 1mim 1t3f 3qyc 4dqo 3c08 3qq9 1jpt 1g9m 1c5c 3k2u 3mxw 3skj 4hk0

H sequence

L sequence

1

Template Search

PIGS
Best H template
Best L template
1-2 template(s)

YES

H or L chain seq. identities < 70%

NO

PIGS
Same antibody for H and L
1 template

MOE
CDR-H3 loop
1 template

Selected template sequences

Selected template structures

2

HMM MSA

IgH alignment

Igκ/Igλ alignment

Combined alignment

3

Model Building

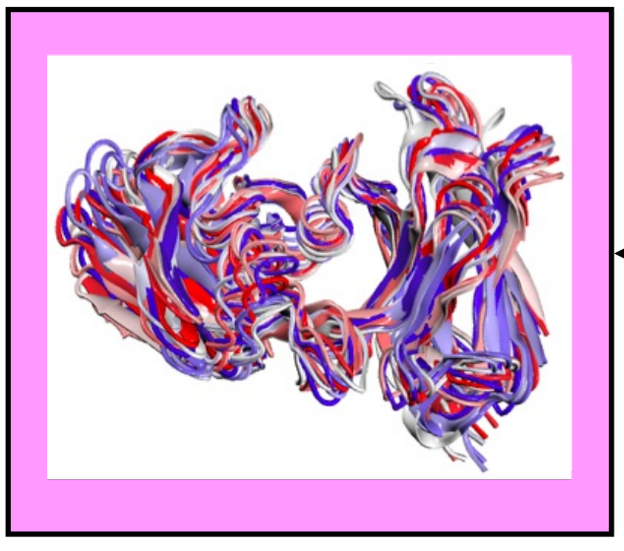
MODELLER
200 models built

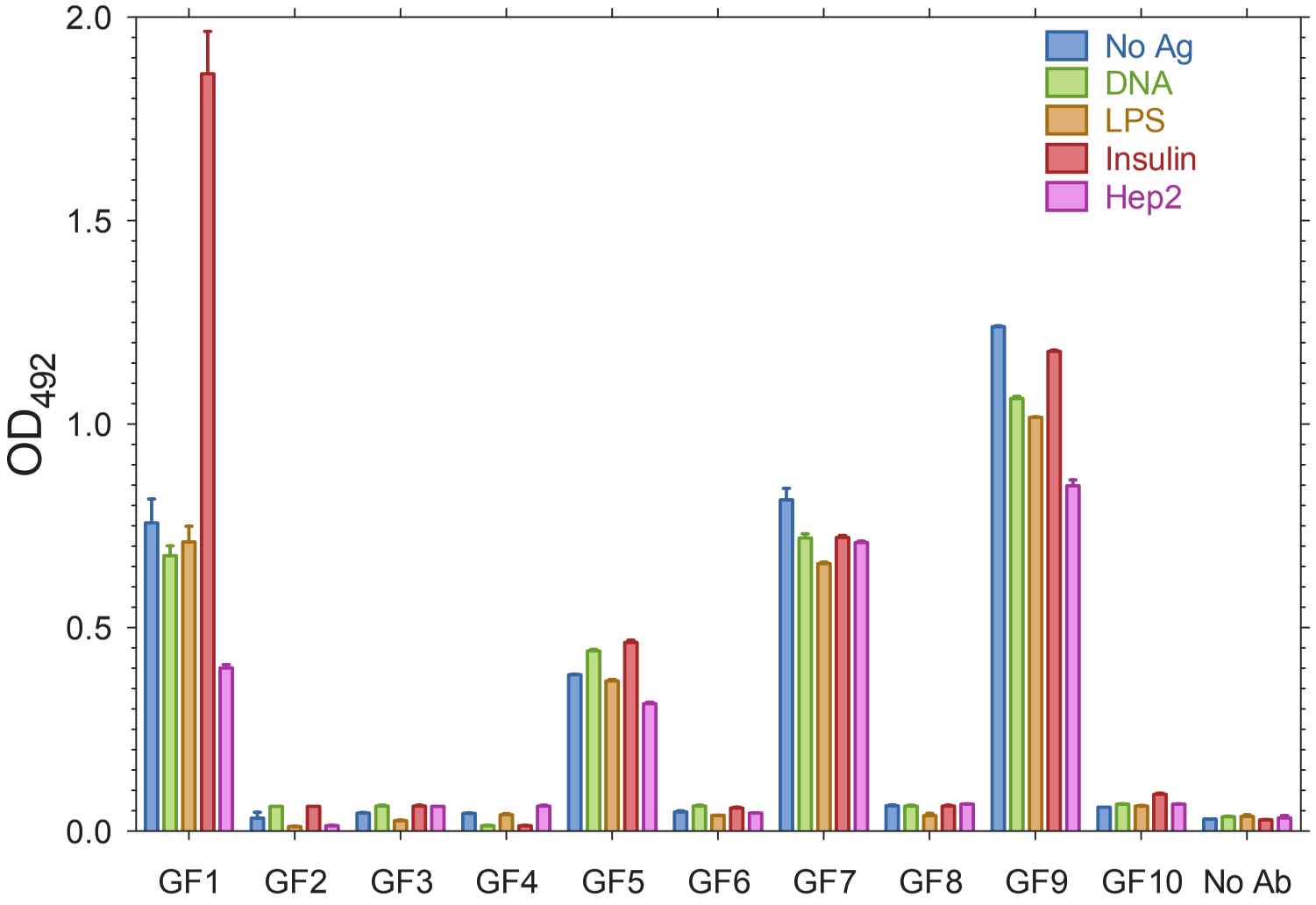
DOPE score
Best model selected (≤ -1)

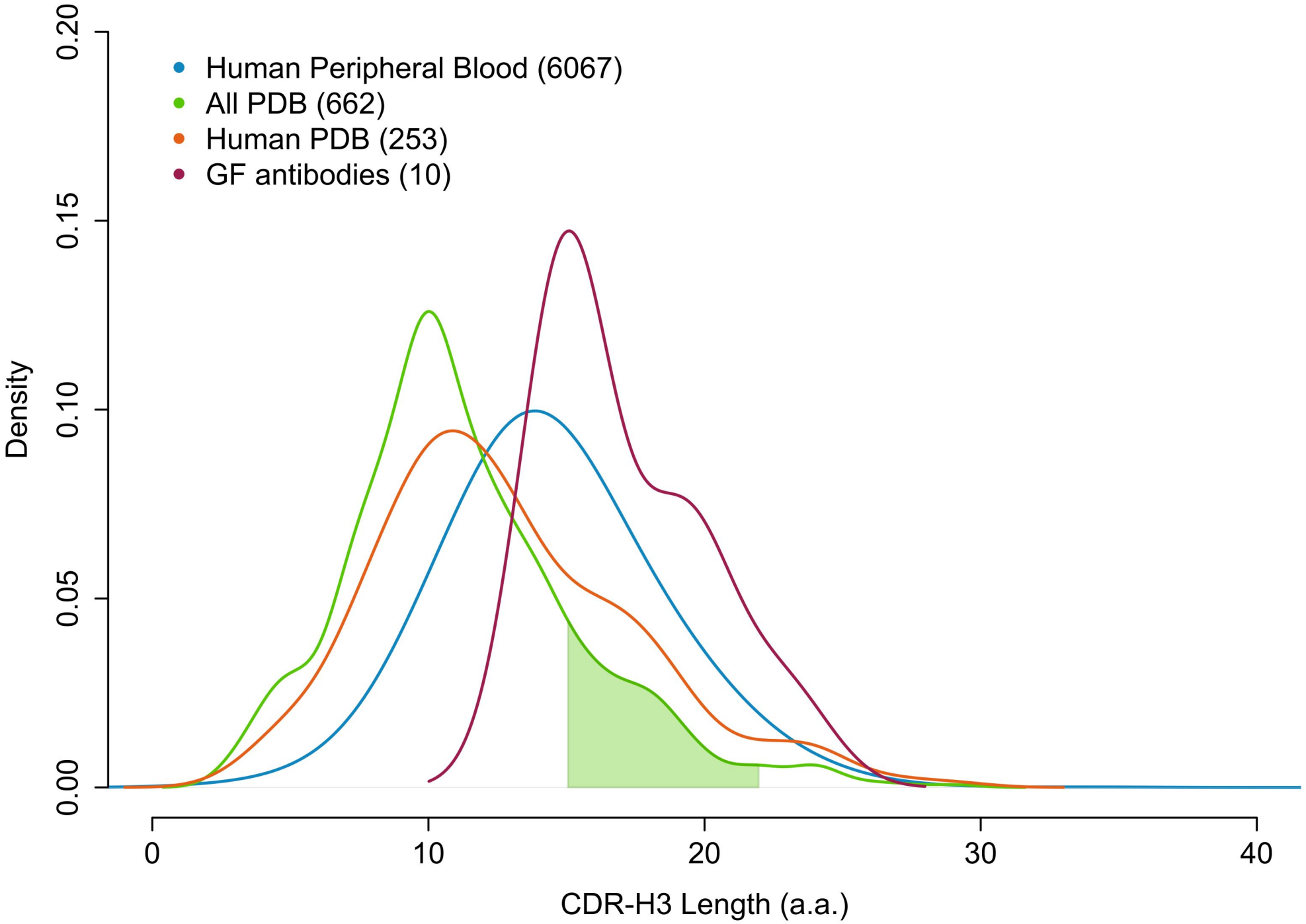
4

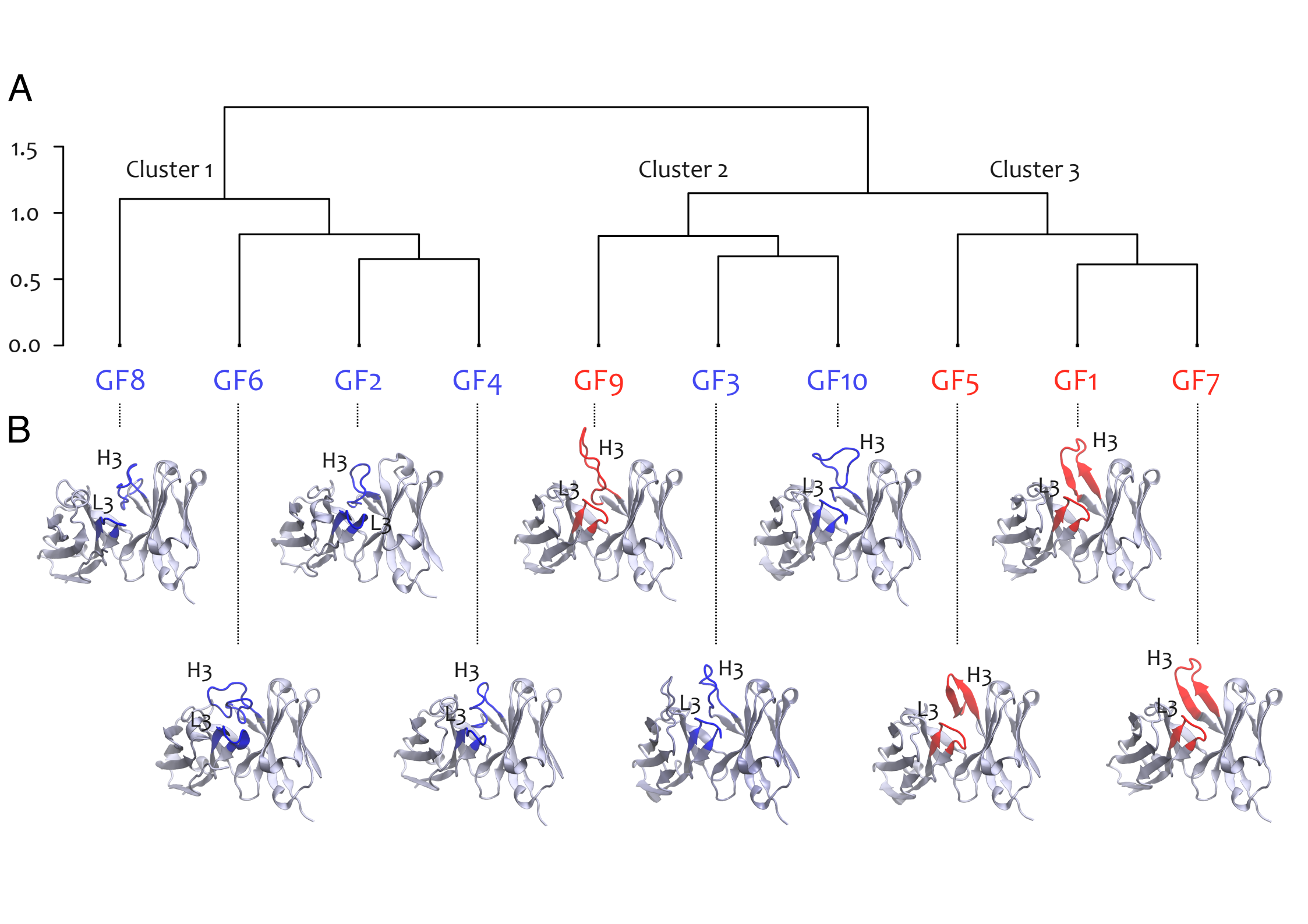
Conformational Sampling

tCONCOORD
500 structures




2.0
1.5
1.0
0.5
0.0
OD492
No Ag
DNA
LPS
Insulin
Hep2
GF1
GF2
GF3
GF4
GF5
GF6
GF7
GF8
GF9
GF10
No Ab

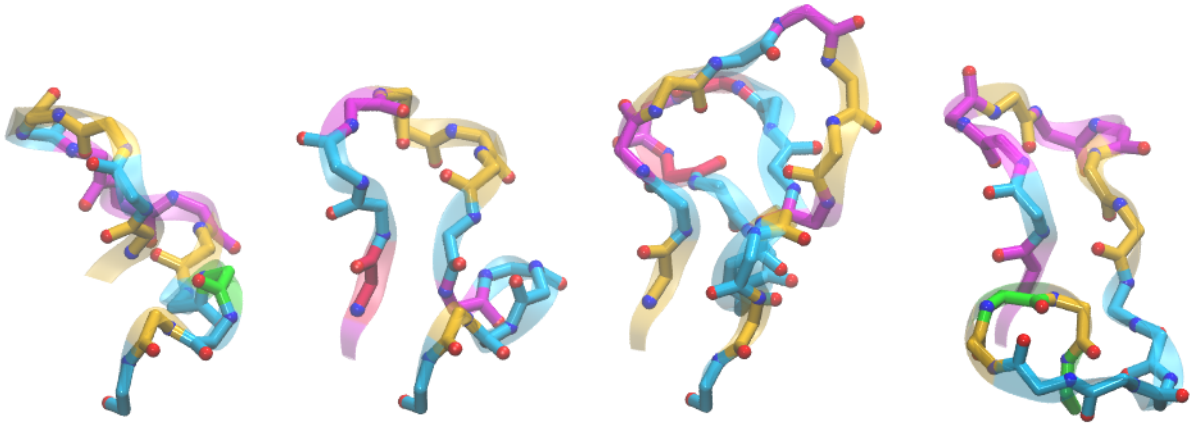

Human Peripheral Blood (6067)
All PDB (662)
Human PDB (253)
GF antibodies (10)
Density
0.00
0.05
0.10
0.15
0.20
0
10
20
30
40
CDR-H3 Length (a.a.)


A
1.5
1.0
0.5
0.0
Cluster 1
Cluster 2
Cluster 3
GF8
GF6
GF2
GF4
GF9
GF3
GF10
GF5
GF1
GF7
B
H3
L3

strong former

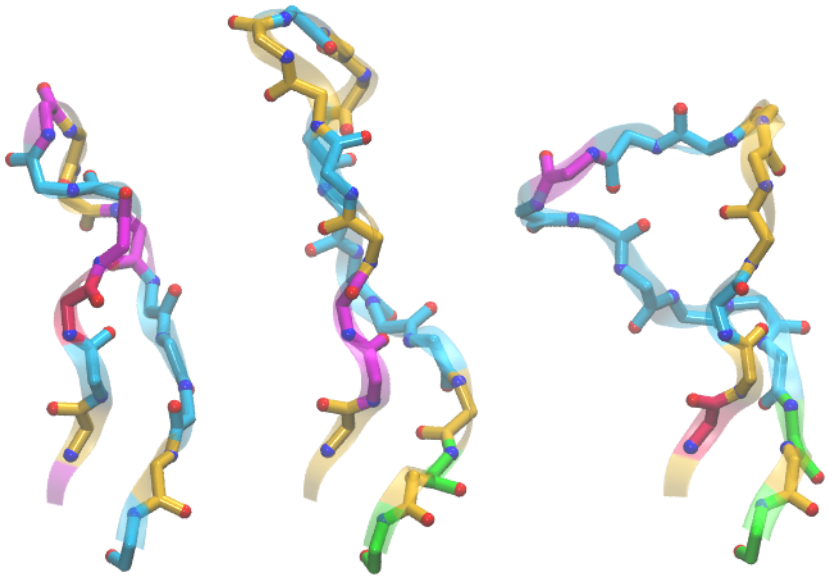
strong breaker

Cluster 1



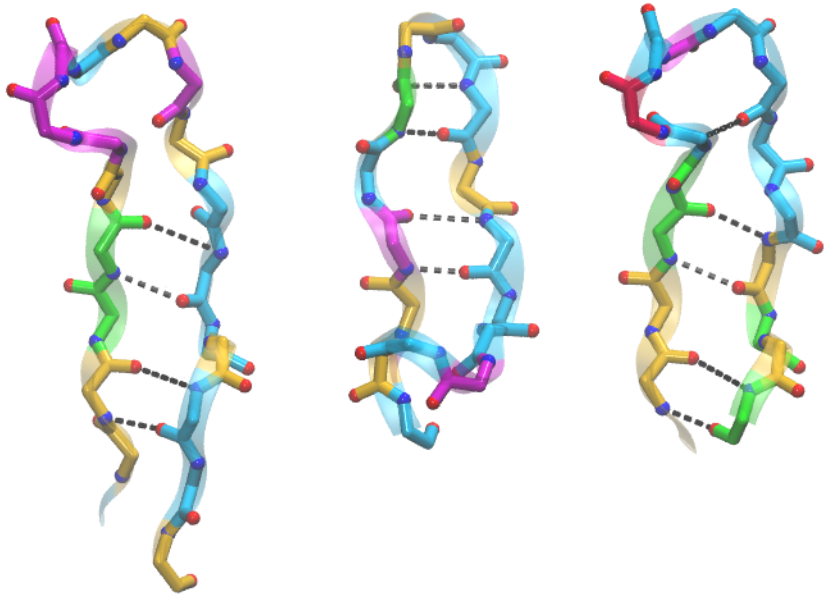
GF2 GF4 GF6 GF8

Cluster 2

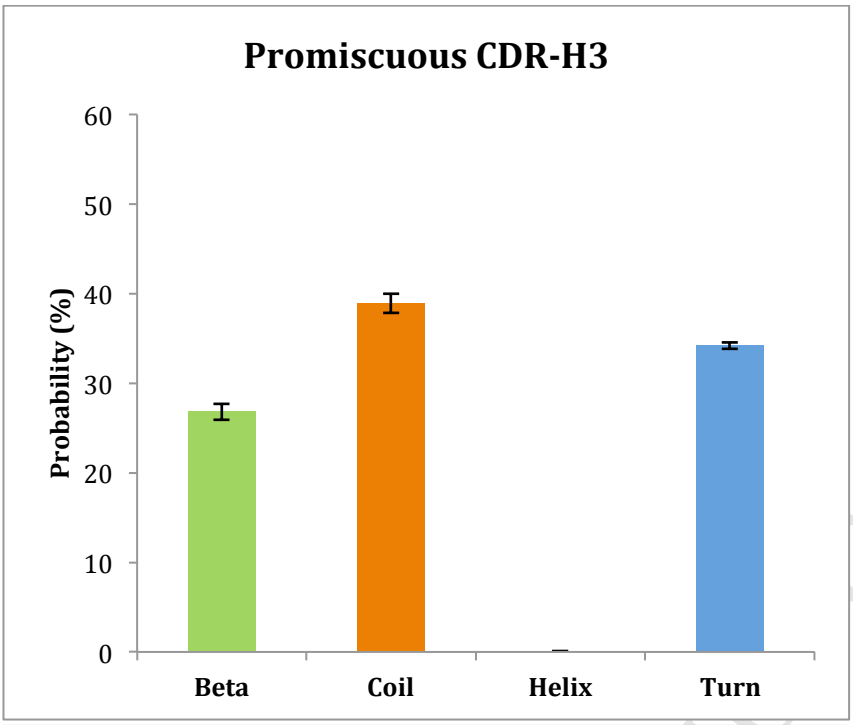


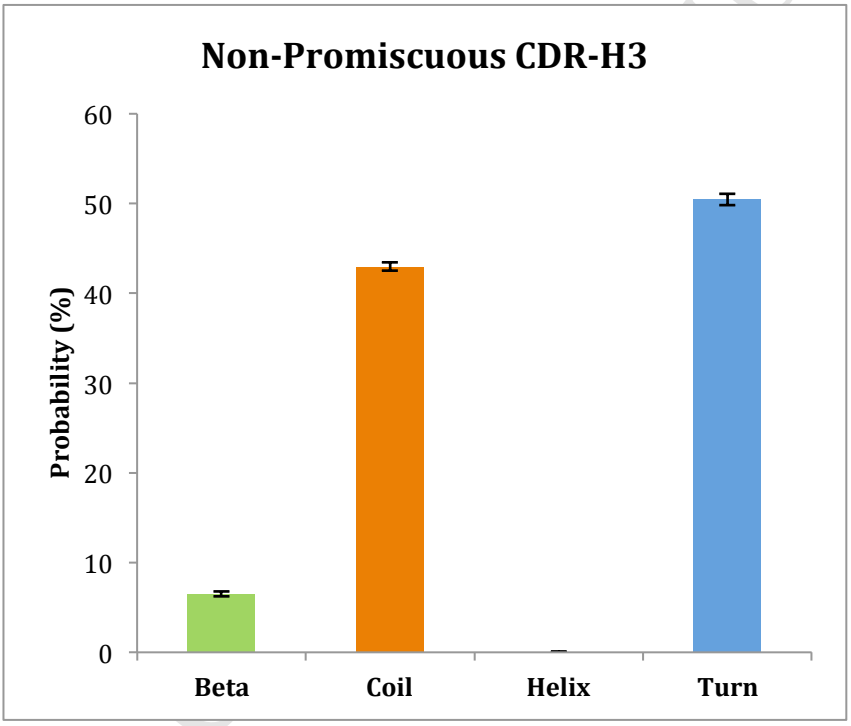
GF3 GF9 GF10

Cluster 3

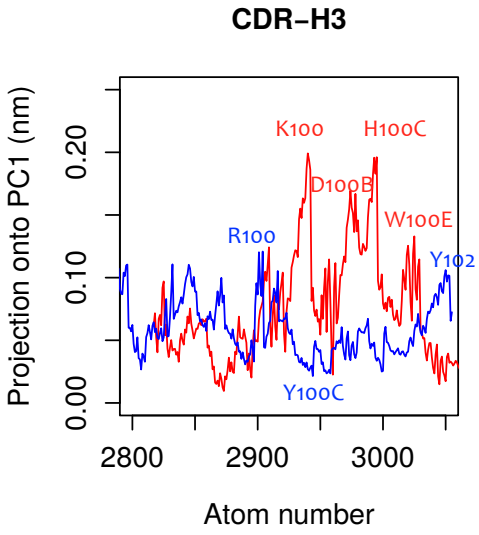


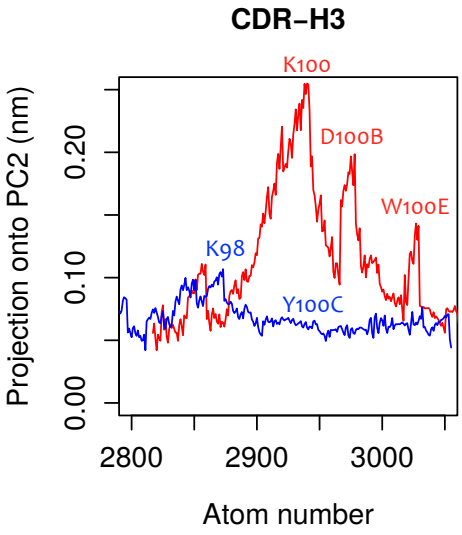
GF1 GF5 GF7


Promiscuous CDR-H3
Probability (%)
60
50
40
30
20
10
0
Beta
Coil
Helix
Turn

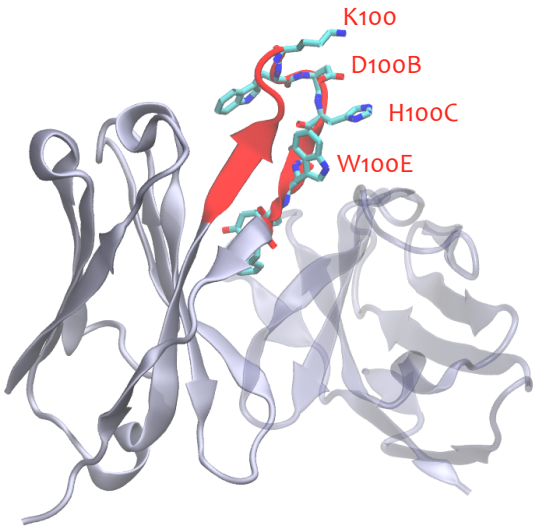

Non-Promiscuous CDR-H3
Probability (%)
60
50
40
30
20
10
0
Beta
Coil
Helix
Turn

A

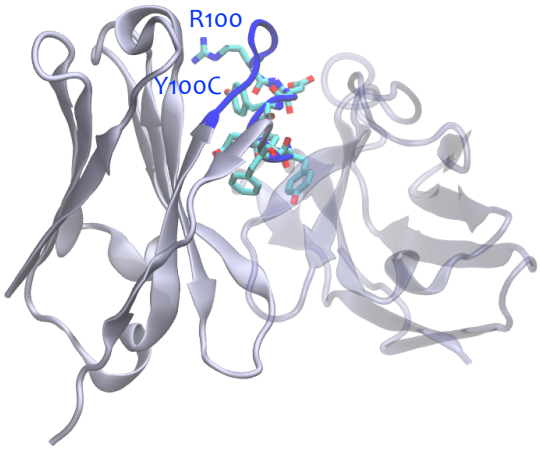

CDR−H3
Projection onto PC1 (nm)
0.00
0.10
0.20
2800
2900
3000
Atom number
K100
H100C
D100B
W100E
R100
Y102
Y100C


CDR−H3
Projection onto PC2 (nm)
0.00
0.10
0.20
2800
2900
3000
Atom number
K100
D100B
W100E
K98
Y100C

B


K100
D100B
H100C
W100E

GF1


R100
Y100C

GF4